VECTORS FOR EXPRESSION OF HML-2 POLYPEPTIDES

All publications and patent applications mentioned in this specification are incorporated herein by reference to the same extent as if each individual document were specifically and individually indicated to be incorporated by reference.

TECHNICAL FIELD

The present invention relates to nucleic acid vectors for polypeptide expression.

BACKGROUND ART

Prostate cancer is the most common type of cancer in men in the USA. Benign prostatic hyperplasia (BPH) is the abnormal growth of benign prostate cells in which the prostate grows and pushes against the urethra and bladder, blocking the normal flow of urine. More than half of the men in the USA aged 60-70 and as many as 90% percent aged 70-90 have symptoms of BPH. Although BPH is seldom a threat to life, it may require treatment to relieve symptoms.

References 1 and 2 disclose that human endogenous retroviruses (HERVs) of the HML-2 subgroup of the HERV-K family show up-regulated expression in prostate tumors. This finding is disclosed as being useful in prostate cancer screening, diagnosis and therapy. In particular, higher levels of an HML-2 expression product relative to normal tissue are said to indicate that the patient from whom the sample was taken has cancer.

Reference 3 discloses that a specific member of the HML-2 family located in chromosome 22 at 20.428 megabases (22q11.2) is preferentially and significantly up-regulated in prostate tumors. This endogenous retrovirus (termed 'PCAV') has several features not found in other members of the HERV-K family: (1) it has a specific nucleotide sequence which distinguishes it from other HERVs within the genome; (2) it has tandem 5' LTRs; (3) it has a fragmented 3' LTR; (4) its *env* gene is interrupted by an alu insertion; and (5) its gag contains a unique insertion. Reference 3 teaches that these features can be exploited in prostate cancer screening, diagnosis and therapy.

References 1 to 3 disclose in general terms vectors for expression of HML-2 and PCAV polypeptides. It is an object of the invention to provide additional and improved vectors for *in vitro* or *in vivo* expression of HML-2 and PCAV polypeptides.

DISCLOSURE OF THE INVENTION

The invention provides a nucleic acid vector comprising: (i) a promoter; (ii) a sequence encoding a HML-2 polypeptide operably linked to said promoter; and (iii) a selectable marker. Preferred vectors further comprise (iv) an origin of replication; and (v) a transcription terminator downstream of and operably linked to (ii).

Vectors of the invention are particularly useful for expression of HML-2 polypeptides either *in vitro* (*e.g.* for later purification) or *in vivo* (*e.g.* for nucleic acid immunization). For use in nucleic acid immunization it is preferred that (i) & (v) should be eukaryotic and (iii) and (iv) should be prokaryotic.

THE PROMOTER

Vectors of the invention include a promoter. It is preferred that the promoter is functional in (*i.e.* can drive transcription in) a eukaryote. The eukaryote is preferably a mammal and more preferably a human. The promoter is preferably active *in vivo*.

The promoter may be a constitutive promoter or it may be a regulated promoter.

The promoter may be specific to particular tissues or cell types, or it may be active in many tissues.

Preferred promoters are viral promoters *e.g.* from cytomegalovirus (CMV). Where viral-based systems are used for delivery, the promoter can be a promoter associated with the respective virus *e.g.* a vaccinia promoter can be used with a vaccinia virus delivery system, *etc.*

The vector may also include transcriptional regulatory sequences (*e.g.* enhancers) in addition to the promoter and which interact functionally with the promoter.

Preferred vectors include the immediate-early CMV enhancer/promoter, and more preferred vectors also include CMV intron A. This was originally isolated from the Towne strain and is very strong. The complete native human immediate-early CMV transcription control unit is divided schematically into four regions from 5' to the ATG of the sequence whose transcription is controlled: I - modulator region (clusters of nuclear factor 1 binding sites); II - enhancers region; III - promoter region; and IV - 5' UTR with intron A. In the native virus, Region I includes upstream sequences that modulate expression in specific cell types and clusters of nuclear factor 1 (NF1) binding sites. Region I can be inhibitory in many cell lines and is generally omitted from vectors of the invention. Regions II and III are generally included in vectors of the invention. Intron A (in Region IV) positively regulates expression in many transformed cell lines and its inclusion enhances expression.

The promoter in vectors of the invention is operably linked to a downstream sequence encoding a HML-2 polypeptide, such that expression of the encoding sequence is under the promoter's control.

THE SEQUENCE ENCODING A HML-2 POLYPEPTIDE

Vectors of the invention include a sequence which encodes a HML-2 polypeptide. The HML-2 is preferably PCAV.

HML-2 is a subgroup of the HERV-K family [4]. HERV isolates which are members of the HML-2 subgroup include HML-2.HOM [5] (also called ERVK6), HERV-K10 [6,7], HERV-K108 [8], the 27 HML-2 viruses shown in Figure 4 of reference 9, HERV-K(C7) [10], HERV-K(II) [11], HERV-K(CH) [1,2]. Because HML-2 is a well-recognized family, the skilled person will be able to determine without difficulty whether any particular HERV-K is or is not a HML-2 *e.g.* by reference to the HERVd database [12].

It is preferred to use sequences from HML-2.HOM, located on chromosome 7 [5, 13], or PCAV [3]. PCAV is a member of the HERV-K sub-family HML2.0, and SEQ ID 75 is the 12366bp sequence of PCAV, based on available human chromosome 22 sequence [14], from the beginning of its first 5' LTR to the end of its fragmented 3' LTR. It is the sense strand of the double-stranded genomic DNA. The transcription start site seems to be at nucleotide 635+5, and its poly-adenylation site is at nucleotide 11735.

The HML-2 polypeptide may be from the gag, prt, pol, env, or cORF regions. HML-2 transcripts which encode these polypeptides are generated by alternative splicing of the full-length mRNA copy of the endogenous viral genome [*e.g.* Figure 4 of ref. 15, Figure 1A of ref. 16, Figure 9 herein]. Although some HML-2 viruses encode all five polypeptides (*e.g.* ERVK6 [5]), the coding regions of most contain mutations which result in one or more coding regions being either mutated or absent. Thus not all HML-2 HERVs have the ability to encode all five polypeptides.

HML-2 gag polypeptide is encoded by the first long ORF in a complete HML-2 genome [17]. Full-length gag polypeptide is proteolytically cleaved. Examples of gag nucleotide sequences are: SEQ ID 1 (HERV-K108); SEQ ID 2 (HERV-K(C7)); SEQ ID 3 (HERV-K(II)); SEQ ID 4 (HERV-K10); and SEQ ID 76 (PCAV). Examples of gag polypeptide sequences are: SEQ ID 5 (HERV-K(C7)); SEQ ID 6 (HERV-K(II)); SEQ IDs 7 & 8 (HERV-K10) ; SEQ ID 9 ('ERVK6'); SEQ ID 69; and SEQ ID 78 (PCAV).

HML-2 prt polypeptide is encoded by the second long ORF in a complete HML-2 genome. It is translated as a gag-prt fusion polypeptide. The fusion polypeptide is proteolytically cleaved to give a protease. Examples of prt nucleotide sequences are: SEQ ID 10 [HERV-K(108)]; SEQ

ID 11 [HERV-K(II)]; SEQ ID 12 [HERV-K10]. Examples of prt polypeptide sequences are: SEQ ID 13 [HERV-K10]; SEQ ID 14 ['ERVK6']; SEQ ID 71.

HML-2 pol polypeptide is encoded by the third long ORF in a complete HML-2 genome. It is translated as a gag-prt-pol fusion polypeptide. The fusion polypeptide is proteolytically cleaved to give three pol products — reverse transcriptase, endonuclease and integrase [18]. Examples of pol nucleotide sequences are: SEQ ID 15 [HERV-K(108)]; SEQ ID 16 [HERV-K(C7)]; SEQ ID 17 [HERV-K(II)]; SEQ ID 18 [HERV-K10]. Examples of pol polypeptide sequences are: SEQ ID 19 [HERV-K(C7)]; SEQ ID 20 [HERV-K10]; SEQ ID 21 ['ERVK6']; SEQ ID 73.

HML-2 env polypeptide is encoded by the fourth long ORF in a complete HML-2 genome. The translated polypeptide is proteolytically cleaved. Examples of env nucleotide sequences are: SEQ ID 22 [HERV-K(108)]; SEQ ID 23 [HERV-K(C7)]; SEQ ID 24 [HERV-K(II)]; SEQ ID 25 [HERV-K10]. Examples of env polypeptide sequences are: SEQ ID 26 [HERV-K(C7)]; SEQ ID 27 [HERV-K10] ; SEQ ID 28 ['ERVK6'].

HML-2 cORF polypeptide is encoded by an ORF which shares the same 5' region and start codon as env. After around 87 codons, a splicing event removes env-coding sequences and the cORF-coding sequence continues in the reading frame +1 relative to that of env [19, 20]. cORF has also been called Rec [21]. Examples of cORF nucleotide sequences are: SEQ IDs 29 & 30 [HERV-K(108)]. An example of a cORF polypeptide sequence is SEQ ID 31.

The HML-2 polypeptide may alternatively be from a PCAP open-reading frame [22], such as PCAP1, PCAP2, PCAP3, PCAP4, PCAP4a or PCAP5 (SEQ IDs 32 to 37 herein). PCAP3 (SEQ IDs 34 & 46) and PCAP5 are preferred (SEQ ID 37).

The HML-2 polypeptide may alternatively be one of SEQ IDs 38 to 50 [22].

Sequences encoding any HML-2 polypeptide expression product may be used in accordance with the invention (*e.g.* sequences encoding any one of SEQ IDs 5, 6, 7, 8, 9, 13, 14, 19, 20, 21, 26, 27, 28, 31-50, 69-74, 78 or 79).

The invention may also utilize sequences encoding polypeptides having at least *a*% identity to such wild-type HML-2 polypeptide sequences. The value of *a* may be 65 or more (*e.g.* 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9). These sequences include allelic variants, SNP variants, homologs, orthologs, paralogs, mutants *etc.* of the SEQ IDs listed in the previous paragraph.

The invention may also utilize sequences having at least *b*% identity to wild-type HML-2 nucleotide sequences. The value of *b* may be 65 or more (*e.g.* 66, 67, 68, 69, 70, 71, 72, 73, 74,

75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9). These sequences include allelic variants, SNP variants, homologs, orthologs, paralogs, mutants *etc.* of SEQ IDs 1, 2, 3, 4, 10, 11, 12, 15, 16, 17, 18, 22, 23, 24, 25, 29 and 30.

The invention may also utilize sequences comprising a fragment of at least *c* nucleotides of such wild-type HML-2 nucleotide sequences. The value of *c* may be 7 or more (*e.g.* 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100, 125, 150, 175, 200, 250, 300 or more). The fragment is preferably a proteolytic cleavage product of a HML-2 polyprotein. The fragment preferably comprises a sequence encoding a T-cell or, preferably, a B-cell epitope from HML-2. T- and B-cell epitopes can be identified empirically (*e.g.* using the PEPSCAN method [23, 24] or similar methods), or they can be predicted *e.g.* using the Jameson-Wolf antigenic index [25], matrix-based approaches [26], TEPITOPE [27], neural networks [28], OptiMer & EpiMer [29, 30], ADEPT [31], Tsites [32], hydrophilicity [33], antigenic index [34] or the methods disclosed in reference 35 *etc.*

The invention may also utilize sequences encoding a polypeptide which comprises a fragment of at least *d* amino acids of wild-type HML-2 polypeptide sequences. The value of *d* may be 7 or more (*e.g.* 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100, 125, 150, 175, 200, 250, 300 or more). The fragment preferably comprises a T-cell or, preferably, a B-cell epitope from HML-2.

The invention may also utilize sequences comprising (i) a first sequence which is a wild-type HML-2 sequence or a sequence as disclosed above and (ii) a second non-HML-2 sequence. Examples of (ii) include sequences encoding: signal peptides, protease cleavage sites, epitopes, leader sequences, tags, fusion partners, N-terminal methionine, arbitrary sequences *etc.* Sequence (ii) will generally be located at the N- and/or C-terminus of (i).

Even though a nucleotide sequence may encode a HML-2 polypeptide which is found naturally, it may differ from the corresponding natural nucleotide sequence. For example, the nucleotide sequence may include mutations *e.g.* to take into account codon preference in a host of interest, or to add restriction sites or tag sequences.

THE SELECTABLE MARKER

Vectors of the invention include a selectable marker.

The marker preferably functions in a microbial host (*e.g.* in a prokaryote, in a bacteria, in a yeast). The marker is preferably a prokaryotic selectable marker (*e.g.* transcribed under the control of a prokaryotic promoter).

For convenience, typical markers are antibiotic resistance genes.

FURTHER FEATURES OF NUCLEIC ACID VECTORS OF THE INVENTION

The vector of the invention is preferably an autonomously replicating episomal or extrachromosomal vector, such as a plasmid.

The vector of the invention preferably comprises an origin of replication. It is preferred that the origin of replication is active in prokaryotes but not in eukaryotes.

Preferred vectors thus include a prokaryotic marker for selection of the vector, a prokaryotic origin of replication, but a *eukaryotic* promoter for driving transcription of the HML-2 coding sequence. The vectors will therefore (a) be amplified and selected in prokaryotic hosts without HML-2 polypeptide expression, but (b) be expressed in eukaryotic hosts without being amplified. This is ideal for nucleic acid immunization vectors.

The vector of the invention may comprise a eukaryotic transcriptional terminator sequence downstream of the HML2-coding sequence. This can enhance transcription levels. Where the HML2-coding sequence does not have its own, the vector of the invention preferably comprises a polyadenylation sequence. A preferred polyadenylation sequence is from bovine growth hormone.

The vector of the invention may comprise a multiple cloning site

In addition to sequences encoding a HML-2 polypeptide and a marker, the vector may comprise a second eukaryotic coding sequence. The vector may also comprise an IRES upstream of said second sequence in order to permit translation of a second eukaryotic polypeptide from the same transcript as the HML-2 polypeptide. Alternatively, the HML-2 polypeptide may be downstream of an IRES.

The vector of the invention may comprise unmethylated CpG motifs *e.g.* unmethylated DNA sequences which have in common a cytosine preceding a guanosine, flanked by two 5' purines and two 3' pyrimidines. In their unmethylated form these DNA motifs have been demonstrated to be potent stimulators of several types of immune cell.

PHARMACEUTICAL COMPOSITIONS

The invention provides a pharmaceutical composition comprising a vector of the invention. The invention also provides the vectors' use as medicaments, and their use in the manufacture of medicaments for treating prostate cancer. The invention also provides a method for treating a patient with a prostate tumor, comprising administering to them a pharmaceutical composition of the invention. The patient is generally a human, preferably a human male, and more preferably an adult human male. Other diseases in which HERV-Ks have been implicated include testicular cancer [36], multiple sclerosis [37], and insulin-dependent diabetes mellitus (IDDM) [38], and the vectors may also be used against these diseases.

The invention also provides a method for raising an immune response, comprising administering an immunogenic dose of a vector of the invention to an animal (*e.g.* to a human).

Pharmaceutical compositions encompassed by the present invention include as active agent, the vectors of the invention in a therapeutically effective amount. An "effective amount" is an amount sufficient to effect beneficial or desired results, including clinical results. An effective amount can be administered in one or more administrations. For purposes of this invention, an effective amount is an amount that is sufficient to palliate, ameliorate, stabilize, reverse, slow or delay the symptoms and/or progression of prostate cancer. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms.

The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. The effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01mg/kg to about 5 mg/kg, or about 0.01 mg/ kg to about 50 mg/kg or about 0.05 mg/kg to about 10 mg/kg of the compositions of the present invention in the individual to which it is administered.

The compositions can be used to treat cancer as well as metastases of primary cancer. In addition, the pharmaceutical compositions can be used in conjunction with conventional methods of cancer treatment, *e.g.* to sensitize tumors to radiation or conventional chemotherapy. The terms "treatment", "treating", "treat" and the like are used herein to generally refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete stabilization or cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, particularly a human, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease symptom, *i.e.* arresting its development; or (c) relieving the disease symptom, *i.e.* causing regression of the disease or symptom.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins,

polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, e.g. mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in reference 39.

The composition is preferably sterile and/or pyrogen-free. It will typically be buffered at about pH 7.

Once formulated, the compositions contemplated by the invention can be (1) administered directly to the subject; or (2) delivered *ex vivo*, to cells derived from the subject (*e.g.* as in *ex vivo* gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, *e.g.* subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

Intramuscular injection is preferred.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art [*e.g.* ref. 40]. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both ex vivo and in vitro applications can be accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the nucleic acid(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Targeted delivery

Vectors of the invention may be delivered in a targeted way.

Receptor-mediated DNA delivery techniques are described in, for example, references 41 to 46. Therapeutic compositions containing a nucleic acid are administered in a range of about 100ng to about 200mg of DNA for local administration in a gene therapy protocol.

Concentration ranges of about 500 ng to about 50 mg, about 1μg to about 2 mg, about 5μg to about 500μg, and about 20μg to about 100μg of DNA can also be used during a gene therapy protocol. Factors such as method of action (*e.g.* for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will affect the dosage required for ultimate efficacy. Where greater expression is desired over a larger area of tissue, larger amounts of vector or the same amounts re-administered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of *e.g.* a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

Vectors can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally references 47 to 50).

Viral-based vectors for delivery of a desired nucleic acid and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (*e.g.* references 51 to 61), alphavirus-based vectors (*e.g.* Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532); hybrids or chimeras of these viruses may also be used), poxvirus vectors (*e.g.* vaccinia, fowlpox, canarypox, modified vaccinia Ankara, *etc.*), adenovirus vectors, and adeno-associated virus (AAV) vectors (*e.g.* see refs. 62 to 67). Administration of DNA linked to killed adenovirus [68] can also be employed.

Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone [*e.g.* 68], ligand-linked DNA [69], eukaryotic cell delivery vehicles cells [*e.g.* refs. 70 to 74] and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in refs. 75 and 76. Liposomes (*e.g.* immunoliposomes) that can act as gene delivery vehicles are described in refs. 77 to 81. Additional approaches are described in refs. 82 & 83.

Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in ref. 83. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials or use of ionizing radiation [*e.g.* refs. 84 & 85]. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun [86] or use of ionizing radiation for activating transferred genes [84 & 87].

Delivery DNA using PLG {poly(lactide-co-glycolide)} microparticles is a particularly preferred method *e.g.* by adsorption to the microparticles, which are optionally treated to have a

negatively-charged surface (*e.g.* treated with SDS) or a positively-charged surface (*e.g.* treated with a cationic detergent, such as CTAB).

Vaccine compositions

The pharmaceutical composition is preferably an immunogenic composition and is more preferably a vaccine composition. Such compositions can be used to raise antibodies in a mammal (*e.g.* a human) and/or to raise a cellular immune response (*e.g.* a response involving T-cells such as CTLs, a response involving natural killer cells, a response involving macrophages *etc.*)

The invention provides the use of a vector of the invention in the manufacture of medicaments for preventing prostate cancer. The invention also provides a method for protecting a patient from prostate cancer, comprising administering to them a pharmaceutical composition of the invention.

Nucleic acid immunization is well known [*e.g.* refs. 88 to 94 *etc.*]

The composition may additionally comprise an adjuvant. For example, the composition may comprise one or more of the following adjuvants: (1) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ [95; Chapter 10 in ref. 96], containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing MTP-PE) formulated into submicron particles using a microfluidizer, (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (2) saponin adjuvants, such as QS21 or StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes), which ISCOMS may be devoid of additional detergent [97]; (3) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (4) cytokines, such as interleukins (*e.g.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 *etc.*), interferons (*e.g.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (5) monophosphoryl lipid A (MPL) or 3-O-deacylated MPL (3dMPL) [*e.g.* 98, 99]; (6) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions [*e.g.* 100, 101, 102]; (7) oligonucleotides comprising CpG motifs *i.e.* containing at least one CG dinucleotide, with 5-methylcytosine optionally being used in place of cytosine; (8) a polyoxyethylene ether or a polyoxyethylene ester [103]; (9) a polyoxyethylene sorbitan ester

surfactant in combination with an octoxynol [104] or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol [105]; (10) an immunostimulatory oligonucleotide (*e.g.* a CpG oligonucleotide) and a saponin [106]; (11) an immunostimulant and a particle of metal salt [107]; (12) a saponin and an oil-in-water emulsion [108]; (13) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) [109]; (14) aluminium salts, preferably hydroxide or phosphate, but any other suitable salt may also be used (*e.g.* hydroxyphosphate, oxyhydroxide, orthophosphate, sulphate etc. [chapters 8 & 9 of ref. 96]). Mixtures of different aluminium salts may also be used. The salt may take any suitable form (*e.g.* gel, crystalline, amorphous etc.); (15) chitosan; (16) cholera toxin or *E.coli* heat labile toxin, or detoxified mutants thereof [110]; (17) microparticles (*i.e.* a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non-toxic (*e.g.* a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone *etc.*, such as poly(lactide-co-glycolide) *etc.*) optionally treated to have a negatively-charged surface (*e.g.* with SDS) or a positively-charged surface (*e.g.* with a cationic detergent, such as CTAB); (18) monophosphoryl lipid A mimics, such as aminoalkyl glucosaminide phosphate derivatives *e.g.* RC-529 [111]; (19) polyphosphazene (PCPP); (20) a bioadhesive [112] such as esterified hyaluronic acid microspheres [113] or a mucoadhesive selected from the group consisting of cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides and carboxymethylcellulose; (21) double-stranded RNA; or (22) other substances that act as immunostimulating agents to enhance the efficacy of the composition. Aluminium salts and/or MF59™ are preferred.

Vaccines of the invention may be prophylactic (*i.e.* to prevent disease) or therapeutic (*i.e.* to reduce or eliminate the symptoms of a disease).

SPECIFIC VECTORS OF THE INVENTION

Preferred vectors of the invention comprise: (i) a eukaryotic promoter; (ii) a sequence encoding a HML-2 polypeptide downstream of and operably linked to said promoter; (iii) a prokaryotic selectable marker; (iv) a prokaryotic origin of replication; and (v) a eukaryotic transcription terminator downstream of and operably linked to said sequence encoding a HML-2 polypeptide.

Particularly preferred vectors are shown in figures 2 to 8 (SEQ IDs 51 to 56 & 80).

VIRUS-LIKE PARTICLES

HML-2 gag polypeptide has been found to assemble into virus-like particles (VLPs). This particulate form of the polypeptide has enhanced immunogenicity when compared to soluble polypeptide and is a preferred form of polypeptide for use in immunization and/or diagnosis.

Thus the invention provides a virus-like particle, comprising HML-2 gag polypeptide. The gag polypeptide may be myristoylated at its N-terminus.

The invention also provides a VLP of the invention for use as an immunogen or for use as a diagnostic antigen. The invention also provides the use of a VLP of the invention in the manufacture of a medicament for immunizing an animal.

The invention also provides a method of raising an immune response in an animal, comprising administering to the animal a VLP of the invention. The immune response may comprise a humoral immune response and/or a cellular immune response.

For raising an immune response, the VLP may be administered with or without an adjuvant as disclosed above. The immune response may treat or protect against cancer (*e.g.* prostate cancer).

The invention also provides a method for diagnosing cancer (*e.g.* prostate cancer) in a patient, comprising the step of contacting antibodies from the patient with VLPs of the invention. Similarly, the invention provides a method for diagnosing cancer (*e.g.* prostate cancer) in a patient, comprising the step of contacting anti-VLP antibodies with a patient sample.

The invention also provides a process for preparing VLPs of the invention, comprising the step of expressing gag polypeptide in a cell, and collecting VLPs from the cell. Expression may be achieved using a vector of the invention.

The VLP of the invention may or may not include packaged nucleic acid.

The gag polypeptide from which the VLPs are made can be from any suitable HML-2 virus (*e.g.* SEQ IDs 1-9, 69 & 78).

DEFINITIONS

The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "about" in relation to a numerical value *x* means, for example, $x \pm 10\%$.

The terms "neoplastic cells", "neoplasia", "tumor", "tumor cells", "cancer" and "cancer cells" (used interchangeably) refer to cells which exhibit relatively autonomous growth, so that they exhibit an aberrant growth phenotype characterized by a significant loss of control of cell

proliferation (*i.e.* de-regulated cell division). Neoplastic cells can be malignant or benign and include prostate cancer derived tissue.

References to a percentage sequence identity between two nucleic acid sequences mean that, when aligned, that percentage of bases are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of reference 114. A preferred alignment program is GCG Gap (Genetics Computer Group, Wisconsin, Suite Version 10.1), preferably using default parameters, which are as follows: open gap = 3; extend gap = 1.

References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of reference 114. A preferred alignment is determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. The Smith-Waterman homology search algorithm is taught in reference 115.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the pCMVkm2 vector, and Figures 2 to 8 show vectors formed by inserting sequences encoding HML-2 polypeptides into this vector.

Figure 9 shows the location of coding sequences in the HML2.HOM genome, with nucleotide numbering according to ref. 5.

Figure 10 is a western blot showing gag expression in transfected 293 cells. Lanes 1 to 4 are: (1) gag opt HML-2; (2) gag opt PCAV; (3) gag wt PCAV; (4) mock.

Figure 11 also shows western blots of transfected 293 cells. In Figure 11A the staining antibody was anti-HML-2, but in Figure 11B it was anti-PCAV. In both 11A and 11B lanes 1 to 4 are: (1) mock; (2) gag opt HML-2; (3) gag opt PCAV; (4) gag wt PCAV. The upper arrow shows the position of gag; the lower arrow shows the β-actin control.

Figure 12 shows electron microscopy of 293 cells expressing (12A) gag opt PCAV or (12B) gag opt HML-2.

MODES FOR CARRYING OUT THE INVENTION

Certain aspects of the present invention are described in greater detail in the non-limiting examples that follow. The examples are put forth so as to provide those of ordinary skill in the art with a disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended

to represent that the experiments below are all and only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.* amounts, temperature, *etc.*) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric.

Vectors for expressing HML-2 polypeptides

The basic pCMVkm2 vector is shown in figure 1. This vector has an immediate-early CMV enhancer/promoter and a bovine growth hormone transcription terminator, with a multiple cloning site in between. The vector also has a kanamycin resistance gene and a ColE1 origin of replication.

Sequences coding for HML-2 polypeptides being inserted between *Sal*I and *Eco*RI in the multiple cloning site:

Figure	SEQ ID	HML-2 polypeptide
2	51	cORF
3	52	PCAP5
4	53	gag
5	54	gag
6	55	Prt
7	56	Pol

Except for the vector shown in figure 4 (SEQ ID 53), the inserted sequences were manipulated for codon preference, including addition of an optimal stop codon:

cORF manipulation:

Start with SEQ ID 57 (SEQ ID 43); manipulate to SEQ ID 58 (SEQ ID 67):

```
ATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATC cORFwt_hml (1)
ATGAACCCCAGCGAGATGCAGCGCAAGGCCCCCCCCCGCCGCCGCCGCCACC corfopt_hml (1)

GCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGA cORFwt_hml (53)
GCAACCGCGCCCCCCTGACCCACAAGATGAACAAGATGGTGACCAGCGAGGA corfopt_hml (53)

ACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAA cORFwt_hml (105)
GCAGATGAAGCTGCCCAGCACCAAGAAGGCCGAGCCCCCCACCTGGGCCCAG corfopt_hml (105)

CTAAAGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGA cORFwt_hml (157)
CTGAAGAAGCTGACCCAGCTGGCCACCAAGTACCTGGAGAACACCAAGGTGA corfopt_hml (157)

CACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGT cORFwt_hml (209)
CCCAGACCCCCGAGAGCATGCTGCTGGCCGCCCTGATGATCGTGAGCATGGT corfopt_hml (209)

GTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGG cORFwt_hml (261)
GAGCGCCGGCGTGCCCAACAGCAGCGAGGAGACCGCCACCATCGAGAACGGC corfopt_hml (261)

CCA---TGA                                              cORFwt_hml (313)
CCCGCTTAA                                              corfopt_hml (313)
```

PCAP5 manipulation:

Start with SEQ ID 59 (SEQ ID 37); manipulate to SEQ ID 60 (SEQ ID 68):

```
ATGAACCCATCGGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACAT pCAP5wt_hml (1)
ATGAACCCCAGCGAGATGCAGCGCAAGGCCCCCCCCCGCCGCCGCCGCCAC pcap5opt_hml (1)

CGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAA pCAP5wt_hml (52)
CGCAACCGCGCCCCCCTGACCCACAAGATGAACAAGATGGTGACCAGCGAG pcap5opt_hml (52)

GAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCA pCAP5wt_hml (103)
GAGCAGATGAAGCTGCCCAGCACCAAGAAGGCCGAGCCCCCCACCTGGGCC pcap5opt_hml (103)

CAACTAAAGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAG pCAP5wt_hml (154)
CAGCTGAAGAAGCTGACCCAGCTGGCCACCAAGTACCTGGAGAACACCAAG pcap5opt_hml (154)

GTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCA pCAP5wt_hml (205)
GTGACCCAGACCCCCGAGAGCATGCTGCTGGCCGCCCTGATGATCGTGAGC pcap5opt_hml (205)

ATGGTGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGC pCAP5wt_hml (256)
ATGGTGGTGTACCCCACCGCCCCCAAGCGCCAGCGCCCCAGCCGCACCGGC pcap5opt_hml (256)

CATGATGACGATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAA pCAP5wt_hml (307)
CACGACGACGACGGCGGCTTCGTGGAGAAGAAGCGCGGCAAGTGCGGCGAG pcap5opt_hml (307)

AAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACAT pCAP5wt_hml (358)
AAGCAGGAGCGCAGCGACTGCTACTGCGTGTGCGTGGAGCGCAGCCGCCAC pcap5opt_hml (358)

AGGAGACTCCATTTTGTTCTGTAC---TAA                        pCAP5wt_hml (409)
CGCCGCCTGCACTTCGTGCTGTACGCTTAA                        pcap5opt_hml (409)
```

Gag manipulation:

Start with SEQ ID 61 (SEQ ID 69); manipulate to SEQ ID 62 (SEQ ID 70):

```
ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCT gagwt_hml (1)
ATGGGCCAGACCAAGAGCAAGATCAAGAGCAAGTACGCCAGCTACCTGAGCT gagopt_hml (1)

TTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAAGTATCTACAAAAAATCT gagwt_hml (53)
TCATCAAGATCCTGCTGAAGCGCGGCGGCGTGAAGGTGAGCACCAAGAACCT gagopt_hml (53)

AATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAA gagwt_hml (105)
GATCAAGCTGTTCCAGATCATCGAGCAGTTCTGCCCCTGGTTCCCCGAGCAG gagopt_hml (105)

GGAACTTTAGATCTAAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAG gagwt_hml (157)
GGCACCCTGGACCTGAAGGACTGGAAGCGCATCGGCAAGGAGCTGAAGCAGG gagopt_hml (157)

CAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCAT gagwt_hml (209)
CCGGCCGCAAGGGCAACATCATCCCCCTGACCGTGTGGAACGACTGGGCCAT gagopt_hml (209)

TATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCGTTTCAGTT gagwt_hml (261)
CATCAAGGCCGCCCTGGAGCCCTTCCAGACCGAGGAGGACAGCGTGAGCGTG gagopt_hml (261)

TCTGATGCCCCTGGAAGCTGTATAATAGATTGTAATGAAAACACAAGGAAAA gagwt_hml (313)
AGCGACGCCCCCGGCAGCTGCATCATCGACTGCAACGAGAACACCCGCAAGA gagopt_hml (313)

AATCCCAGAAAGAAACGGAAGGTTTACATTGCGAATATGTAGCAGAGCCGGT gagwt_hml (365)
AGAGCCAGAAGGAGACCGAGGGCCTGCACTGCGAGTACGTGGCCGAGCCCGT gagopt_hml (365)

AATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGAGGTGATA gagwt_hml (417)
GATGGCCCAGAGCACCCAGAACGTGGACTACAACCAGCTGCAGGAGGTGATC gagopt_hml (417)

TATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCAT gagwt_hml (469)
TACCCCGAGACCCTGAAGCTGGAGGGCAAGGGCCCCGAGCTGGTGGGCCCCA gagopt_hml (469)

CAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTCAGGTGCCTGT gagwt_hml (521)
GCGAGAGCAAGCCCCGCGGCACCAGCCCCCTGCCCGCCGGCCAGGTGCCCGT gagopt_hml (521)

AACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTA gagwt_hml (573)
```

```
GACCCTGCAGCCCCAGAAGCAGGTGAAGGAGAACAAGACCCAGCCCCCCGTG gagopt_hml (573)

GCCTATCAATACTGGCCTCCGGCTGAACTTCAGTATCGGCCACCCCCAGAAA gagwt_hml (625)
GCCTACCAGTACTGGCCCCCCGCCGAGCTGCAGTACCGCCCCCCCCCCGAGA gagopt_hml (625)

GTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGGGCGCCATA gagwt_hml (677)
GCCAGTACGGCTACCCCGGCATGCCCCCCGCCCCCCAGGGCCGCGCCCCCTA gagopt_hml (677)

CCCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAG gagwt_hml (729)
CCCCCAGCCCCCCACCCGCCGCCTGAACCCCACCGCCCCCCCCAGCCGCCAG gagopt_hml (729)

GGTAGTAAATTACATGAAATTATTGATAAATCAAGAAAGGAAGGAGATACTG gagwt_hml (781)
GGCAGCAAGCTGCACGAGATCATCGACAAGAGCCGCAAGGAGGGCGACACCG gagopt_hml (781)

AGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAGGAGC gagwt_hml (833)
AGGCCTGGCAGTTCCCCGTGACCCTGGAGCCCATGCCCCCCGGCGAGGGCGC gagopt_hml (833)

CCAAGAGGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATA gagwt_hml (885)
CCAGGAGGGCGAGCCCCCCACCGTGGAGGCCCGCTACAAGAGCTTCAGCATC gagopt_hml (885)

AAAAAGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGACCCAACTCCC gagwt_hml (937)
AAGAAGCTGAAGGACATGAAGGAGGGCGTGAAGCAGTACGGCCCCAACAGCC gagopt_hml (937)

CTTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCC gagwt_hml (989)
CCTACATGCGCACCCTGCTGGACAGCATCGCCCACGGCCACCGCCTGATCCC gagopt_hml (989)

TTATGATTGGGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTA gagwt_hml (1041)
CTACGACTGGGAGATCCTGGCCAAGAGCAGCCTGAGCCCCAGCCAGTTCCTG gagopt_hml (1041)

CAATTTAAGACTTGGTGGATTGATGGGGTACAAGAACAGGTCCGAAGAAATA gagwt_hml (1093)
CAGTTCAAGACCTGGTGGATCGACGGCGTGCAGGAGCAGGTGCGCCGCAACC gagopt_hml (1093)

GGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGG gagwt_hml (1145)
GCGCCGCCAACCCCCCCGTGAACATCGACGCCGACCAGCTGCTGGGCATCGG gagopt_hml (1145)

TCAAAATTGGAGTACTATTAGTCAACAAGCATTAATGCAAAATGAGGCCATT gagwt_hml (1197)
CCAGAACTGGAGCACCATCAGCCAGCAGGCCCTGATGCAGAACGAGGCCATC gagopt_hml (1197)

GAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAATCCAAGACCCAG gagwt_hml (1249)
GAGCAGGTGCGCGCCATCTGCCTGCGCGCCTGGGAGAAGATCCAGGACCCCG gagopt_hml (1249)

GAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTA gagwt_hml (1301)
GCAGCACCTGCCCCAGCTTCAACACCGTGCGCCAGGGCAGCAAGGAGCCCTA gagopt_hml (1301)

TCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCTGAT gagwt_hml (1353)
CCCCGACTTCGTGGCCCGCCTGCAGGACGTGGCCCAGAAGAGCATCGCCGAC gagopt_hml (1353)

GAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCATATGAAAACGCCAATC gagwt_hml (1405)
GAGAAGGCCCGCAAGGTGATCGTGGAGCTGATGGCCTACGAGAACGCCAACC gagopt_hml (1405)

CTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATC gagwt_hml (1457)
CCGAGTGCCAGAGCGCCATCAAGCCCCTGAAGGGCAAGGTGCCCGCCGGCAG gagopt_hml (1457)

AGATGTAATCTCAGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATG gagwt_hml (1509)
CGACGTGATCAGCGAGTACGTGAAGGCCTGCGACGGCATCGGCGGCGCCATG gagopt_hml (1509)

CATAAAGCTATGCTTATGGCTCAAGCAATAACAGGAGTTGTTTTAGGAGGAC gagwt_hml (1561)
CACAAGGCCATGCTGATGGCCCAGGCCATCACCGGCGTGGTGCTGGGCGGCC gagopt_hml (1561)

AAGTTAGAACATTTGGAAGAAAATGTTATAATTGTGGTCAAATTGGTCACTT gagwt_hml (1613)
AGGTGCGCACCTTCGGCCGCAAGTGCTACAACTGCGGCCAGATCGGCCACCT gagopt_hml (1613)

AAAAAAGAATTGCCCAGTCTTAAATAAACAGAATATAACTATTCAAGCAACT gagwt_hml (1665)
GAAGAAGAACTGCCCCGTGCTGAACAAGCAGAACATCACCATCCAGGCCACC gagopt_hml (1665)

ACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAAC gagwt_hml (1717)
ACCACCGGCCGCGAGCCCCCCGACCTGTGCCCCCGCTGCAAGAAGGGCAAGC gagopt_hml (1717)

ATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTC gagwt_hml (1769)
ACTGGGCCAGCCAGTGCCGCAGCAAGTTCGACAAGAACGGCCAGCCCCTGAG gagopt_hml (1769)
```

```
GGGAAACGAGCAAAGGGGCCAGCCTCAGGCCCCACAACAAACTGGGGCATTC gagwt_hml (1821)
CGGCAACGAGCAGCGCGGCCAGCCCCAGGCCCCCCAGCAGACCGGCGCCTTC gagopt_hml (1821)

CCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCACTGT gagwt_hml (1873)
CCCATCCAGCCCTTCGTGCCCCAGGGCTTCCAGGGCCAGCAGCCCCCCCTGA gagopt_hml (1873)

CCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCC gagwt_hml (1925)
GCCAGGTGTTCCAGGGCATCAGCCAGCTGCCCCAGTACAACAACTGCCCCCC gagopt_hml (1925)

GCCACAAGCGGCAGTGCAGCAG---TAG                          gagwt_hml (1977)
CCCCCAGGCCGCCGTGCAGCAGGCTTAA                          gagopt_hml (1977)
```

Prt manipulation:

Start with SEQ ID 63 (SEQ ID 71); manipulate to SEQ ID 64 (SEQ ID 72):

```
ATGTGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAGGCCCCA Protwt_hml (1)
ATGTGGGCCACCATCGTGGGCAAGCGCGCCAAGGGCCCCGCCAGCGGCCCCA protopt_hml (1)

CAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGG Protwt_hml (53)
CCACCAACTGGGGCATCCCCAACAGCGCCATCTGCAGCAGCGGCTTCAGCGG protopt_hml (53)

GACAACAACCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACA Protwt_hml (105)
CACCACCACCCCCACCGTGCCCAGCGTGAGCGGCAACAAGCCCGTGACCACC protopt_hml (105)

ATACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGTA Protwt_hml (157)
ATCCAGCAGCTGAGCCCCGCCACCAGCGGCAGCGCCGCCGTGGACCTGTGCA protopt_hml (157)

CTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCCACAAAAAACCCCCAC Protwt_hml (209)
CCATCCAGGCCGTGAGCCTGCTGCCCGGCGAGCCCCCCCAGAAGACCCCCAC protopt_hml (209)

AGGGGTATATGGACCCCTGCCTAAGGGGACTGTAGGACTAATCTTGGGACGA Protwt_hml (261)
CGGCGTGTACGGCCCCCTGCCCAAGGGCACCGTGGGCCTGATCCTGGGCCGC protopt_hml (261)

TCAAGTCTAAATCTAAAAGGAGTTCAAATTCATACTAGTGTGGTTGATTCAG Protwt_hml (313)
AGCAGCCTGAACCTGAAGGGCGTGCAGATCCACACCAGCGTGGTGGACAGCG protopt_hml (313)

ACTATAAAGGCGAAATTCAATTGGTTATTAGCTCTTCAATTCCTTGGAGTGC Protwt_hml (365)
ACTACAAGGGCGAGATCCAGCTGGTGATCAGCAGCAGCATCCCCTGGAGCGC protopt_hml (365)

CAGTCCAAGAGACAGGATTGCTCAATTATTACTCCTGCCATACATTAAGGGT Protwt_hml (417)
CAGCCCCCGCGACCGCATCGCCCAGCTGCTGCTGCTGCCCTACATCAAGGGC protopt_hml (417)

GGAAATAGTGAAATAAAAAGAATAGGAGGGCTTGGAAGCACTGATCCAACAG Protwt_hml (469)
GGCAACAGCGAGATCAAGCGCATCGGCGGCCTGGGCAGCACCGACCCCACCG protopt_hml (469)

GAAAGGCTGCATATTGGGCAAGTCAGGTCTCAGAGAACAGACCTGTGTGTAA Protwt_hml (521)
GCAAGGCCGCCTACTGGGCCAGCCAGGTGAGCGAGAACCGCCCCGTGTGCAA protopt_hml (521)

GGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGGTAGACACTGGAGCAGAT Protwt_hml (573)
GGCCATCATCCAGGGCAAGCAGTTCGAGGGCCTGGTGGACACCGGCGCCGAC protopt_hml (573)

GTCTCTATCATTGCTTTAAATCAGTGGCCAAAAAATTGGCCTAAACAAAAGG Protwt_hml (625)
GTGAGCATCATCGCCCTGAACCAGTGGCCCAAGAACTGGCCCAAGCAGAAGG protopt_hml (625)

CTGTTACAGGACTTGTCGGCATAGGCACAGCCTCAGAAGTGTATCAAAGTAC Protwt_hml (677)
CCGTGACCGGCCTGGTGGGCATCGGCACCGCCAGCGAGGTGTACCAGAGCAC protopt_hml (677)

GGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAGTACTGTTCAGCCA Protwt_hml (729)
CGAGATCCTGCACTGCCTGGGCCCCGACAACCAGGAGAGCACCGTGCAGCCC protopt_hml (729)

ATGATTACTTCAATTCCTCTTAATCTGTGGGGTCGAGATTTATTACAACAAT Protwt_hml (781)
ATGATCACCAGCATCCCCCTGAACCTGTGGGGCCGCGACCTGCTGCAGCAGT protopt_hml (781)

GGGGTGCGGAAATCACCATGCCCGCTCCATCATATAGCCCCACGAGTCAAAA Protwt_hml (833)
GGGGCGCCGAGATCACCATGCCCGCCCCCAGCTACAGCCCCACCAGCCAGAA protopt_hml (833)

AATCATGACCAAGATGGGATATATACCAGGAAAGGGACTAGGGAAAAATGAA Protwt_hml (885)
```

```
GATCATGACCAAGATGGGCTACATCCCCGGCAAGGGCCTGGGCAAGAACGAG protopt_hml (885)

GATGGCATTAAAATTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGAA Protwt_hml (937)
GACGGCATCAAGATCCCCGTGGAGGCCAAGATCAACCAGGAGCGCGAGGGCA protopt_hml (937)

TAGGGAATCCTTGC---TAG                                  Protwt_hml (989)
TCGGCAACCCCTGCGCTTAA                                  protopt_hml (989)
```

Pol manipulation:

Start with SEQ ID 65 (SEQ ID 73); manipulate to SEQ ID 66 (SEQ ID 74):

```
ATGAATAAATCAAGAAAGAGAAGGAATAGGGAATCCTTGCTAGGGGCGGCCA polwt_hml (1)
ATGAACAAGAGCCGCAAGCGCCGCAACCGCGAGAGCCTGCTGGGCGCCGCCA polopt_hml (1)

CTGTAGAGCCTCCTAAACCCATACCATTAACTTGGAAAACAGAAAAACCAGT polwt_hml (53)
CCGTGGAGCCCCCCAAGCCCATCCCCCTGACCTGGAAGACCGAGAAGCCCGT polopt_hml (53)

GTGGGTAAATCAGTGGCCGCTACCAAAACAAAAACTGGAGGCTTTACATTTA polwt_hml (105)
GTGGGTGAACCAGTGGCCCCTGCCCAAGCAGAAGCTGGAGGCCCTGCACCTG polopt_hml (105)

TTAGCAAATGAACAGTTAGAAAAGGGTCATATTGAGCCTTCGTTCTCACCTT polwt_hml (157)
CTGGCCAACGAGCAGCTGGAGAAGGGCCACATCGAGCCCAGCTTCAGCCCCT polopt_hml (157)

GGAATTCTCCTGTGTTTGTAATTCAGAAGAAATCAGGCAAATGGCGTATGTT polwt_hml (209)
GGAACAGCCCCGTGTTCGTGATCCAGAAGAAGAGCGGCAAGTGGCGCATGCT polopt_hml (209)

AACTGACTTAAGGGCTGTAAACGCCGTAATTCAACCCATGGGGCCTCTCCAA polwt_hml (261)
GACCGACCTGCGCGCCGTGAACGCCGTGATCCAGCCCATGGGCCCCCTGCAG polopt_hml (261)

CCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAA polwt_hml (313)
CCCGGCCTGCCCAGCCCCGCCATGATCCCCAAGGACTGGCCCCTGATCATCA polopt_hml (313)

TTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGCGA polwt_hml (365)
TCGACCTGAAGGACTGCTTCTTCACCATCCCCCTGGCCGAGCAGGACTGCGA polopt_hml (365)

AAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGG polwt_hml (417)
GAAGTTCGCCTTCACCATCCCCGCCATCAACAACAAGGAGCCCGCCACCCGC polopt_hml (417)

TTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTC polwt_hml (469)
TTCCAGTGGAAGGTGCTGCCCCAGGGCATGCTGAACAGCCCCACCATCTGCC polopt_hml (469)

AGACTTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTG polwt_hml (521)
AGACCTTCGTGGGCCGCGCCCTGCAGCCCGTGCGCGAGAAGTTCAGCGACTG polopt_hml (521)

TTATATTATTCATTGTATTGATGATATTTTATGTGCTGCAGAAACGAAAGAT polwt_hml (573)
CTACATCATCCACTGCATCGACGACATCCTGTGCGCCGCCGAGACCAAGGAC polopt_hml (573)

AAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGAC polwt_hml (625)
AAGCTGATCGACTGCTACACCTTCCTGCAGGCCGAGGTGGCCAACGCCGGCC polopt_hml (625)

TGGCAATAGCATCTGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGG polwt_hml (677)
TGGCCATCGCCAGCGACAAGATCCAGACCAGCACCCCCTTCCACTACCTGGG polopt_hml (677)

GATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGAAAA polwt_hml (729)
CATGCAGATCGAGAACCGCAAGATCAAGCCCCAGAAGATCGAGATCCGCAAG polopt_hml (729)

GACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATT polwt_hml (781)
GACACCCTGAAGACCCTGAACGACTTCCAGAAGCTGCTGGGCGACATCAACT polopt_hml (781)

GGATTCGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTC polwt_hml (833)
GGATCCGCCCCACCCTGGGCATCCCCACCTACGCCATGAGCAACCTGTTCAG polopt_hml (833)

TATCTTAAGAGGAGACTCAGACTTAAATAGTAAAAGAATGTTAACCCCAGAG polwt_hml (885)
CATCCTGCGCGGCGACAGCGACCTGAACAGCAAGCGCATGCTGACCCCCGAG polopt_hml (885)

GCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAA polwt_hml (937)
GCCACCAAGGAGATCAAGCTGGTGGAGGAGAAGATCCAGAGCGCCCAGATCA polopt_hml (937)
```

```
ATAGAATAGATCCCTTAGCCCCACTCCAACTTTTGATTTTTGCCACTGCACA polwt_hml (989)
ACCGCATCGACCCCCTGGCCCCCCTGCAGCTGCTGATCTTCGCCACCGCCCA polopt_hml (989)

TTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTGGTCATTC polwt_hml (1041)
CAGCCCCACCGGCATCATCATCCAGAACACCGACCTGGTGGAGTGGAGCTTC polopt_hml (1041)

CTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTA polwt_hml (1093)
CTGCCCCACAGCACCGTGAAGACCTTCACCCTGTACCTGGACCAGATCGCCA polopt_hml (1093)

CATTAATCGGTCAGACAAGATTACGAATAATAAAATTATGTGGGAATGACCC polwt_hml (1145)
CCCTGATCGGCCAGACCCGCCTGCGCATCATCAAGCTGTGCGGCAACGACCC polopt_hml (1145)

AGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATC polwt_hml (1197)
CGACAAGATCGTGGTGCCCCTGACCAAGGAGCAGGTGCGCCAGGCCTTCATC polopt_hml (1197)

AATTCTGGTGCATGGAAGATTGGTCTTGCTAATTTTGTGGGAATTATTGATA polwt_hml (1249)
AACAGCGGCGCCTGGAAGATCGGCCTGGCCAACTTCGTGGGCATCATCGACA polopt_hml (1249)

ATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTGACTACTTGGAT polwt_hml (1301)
ACCACTACCCCAAGACCAAGATCTTCCAGTTCCTGAAGCTGACCACCTGGAT polopt_hml (1301)

TCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTT polwt_hml (1353)
CCTGCCCAAGATCACCCGCCGCGAGCCCCTGGAGAACGCCCTGACCGTGTTC polopt_hml (1353)

ACTGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGACCGAAAGAACGAG polwt_hml (1405)
ACCGACGGCAGCAGCAACGGCAAGGCCGCCTACACCGGCCCCAAGGAGCGCG polopt_hml (1405)

TAATCAAAACTCCATATCAATCGGCTCAAAGAGCAGAGTTGGTTGCAGTCAT polwt_hml (1457)
TGATCAAGACCCCCTACCAGAGCGCCCAGCGCGCCGAGCTGGTGGCCGTGAT polopt_hml (1457)

TACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCA polwt_hml (1509)
CACCGTGCTGCAGGACTTCGACCAGCCCATCAACATCATCAGCGACAGCGCC polopt_hml (1509)

TATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCA polwt_hml (1561)
TACGTGGTGCAGGCCACCCGCGACGTGGAGACCGCCCTGATCAAGTACAGCA polopt_hml (1561)

TGGATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAACTGTAAGAAA polwt_hml (1613)
TGGACGACCAGCTGAACCAGCTGTTCAACCTGCTGCAGCAGACCGTGCGCAA polopt_hml (1613)

AAGAAATTTCCCATTTTATATTACACATATTCGAGCACACACTAATTTACCA polwt_hml (1665)
GCGCAACTTCCCCTTCTACATCACCCACATCCGCGCCCACACCAACCTGCCC polopt_hml (1665)

GGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGT-ATCATCTGCA polwt_hml (1717)
GGCCCCCTGACCAAGGCCAACGAGCAGGCCGACCTGCTGGTGAGCAGC-GCC polopt_hml (1717)

CTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCAGGAT polwt_hml (1768)
CTGATCAAGGCCCAGGAGCTGCACGCCCTGACCCACGTGAACGCCGCCGGCC polopt_hml (1768)

TAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACA polwt_hml (1820)
TGAAGAACAAGTTCGACGTGACCTGGAAGCAGGCCAAGGACATCGTGCAGCA polopt_hml (1820)

TTGCACCCAGTGTCAAGTCTTACACCTGCCCACTCAAGAGGCAGGAGTTAAT polwt_hml (1872)
CTGCACCCAGTGCCAGGTGCTGCACCTGCCCACCCAGGAGGCCGGCGTGAAC polopt_hml (1872)

CCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCACGCATGTAC polwt_hml (1924)
CCCCGCGGCCTGTGCCCCAACGCCCTGTGGCAGATGGACGTGACCCACGTGC polopt_hml (1924)

CTTCATTTGGAAGATTATCATATGTTCACGTAACAGTTGATACTTATTCACA polwt_hml (1976)
CCAGCTTCGGCCGCCTGAGCTACGTGCACGTGACCGTGGACACCTACAGCCA polopt_hml (1976)

TTTCATATGGGCAACTTGCCAAACAGGAGAAAGTACTTCCCATGTTAAAAAA polwt_hml (2028)
CTTCATCTGGGCCACCTGCCAGACCGGCGAGAGCACCAGCCACGTGAAGAAG polopt_hml (2028)

CATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTG polwt_hml (2080)
CACCTGCTGAGCTGCTTCGCCGTGATGGGCGTGCCCGAGAAGATCAAGACCG polopt_hml (2080)

ACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTG polwt_hml (2132)
ACAACGGCCCCGGCTACTGCAGCAAGGCCTTCCAGAAGTTCCTGAGCCAGTG polopt_hml (2132)

GAAAATTTCACATACAACAGGAATTCCTTATAATTCCCAAGGACAGGCCATA polwt_hml (2184)
```

```
GAAGATCAGCCACACCACCGGCATCCCCTACAACAGCCAGGGCCAGGCCATC polopt_hml (2184)

GTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAG polwt_hml (2236)
GTGGAGCGCACCAACCGCACCCTGAAGACCCAGCTGGTGAAGCAGAAGGAGG polopt_hml (2236)

GGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACT polwt_hml (2288)
GCGGCGACAGCAAGGAGTGCACCACCCCCCAGATGCAGCTGAACCTGGCCCT polopt_hml (2288)

CTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCA polwt_hml (2340)
GTACACCCTGAACTTCCTGAACATCTACCGCAACCAGACCACCACCAGCGCC polopt_hml (2340)

GAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTT polwt_hml (2392)
GAGCAGCACCTGACCGGCAAGAAGAACAGCCCCCACGAGGGCAAGCTGATCT polopt_hml (2392)

GGTGGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTG polwt_hml (2444)
GGTGGAAGGACAACAAGAACAAGACCTGGGAGATCGGCAAGGTGATCACCTG polopt_hml (2444)

GGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGG polwt_hml (2496)
GGGCCGCGGCTTCGCCTGCGTGAGCCCCGGCGAGAACCAGCTGCCCGTGTGG polopt_hml (2496)

ATACCCACTAGACATTTGAAGTTCTACAATGAACCCATCAGAGATGCAAAGA polwt_hml (2548)
ATCCCCACCCGCCACCTGAAGTTCTACAACGAGCCCATCCGCGACGCCAAGA polopt_hml (2548)

AAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACA polwt_hml (2600)
AGAGCACCAGCGCCGAGACCGAGACCAGCCAGAGCAGCACCGTGGACAGCCA polopt_hml (2600)

AGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAA polwt_hml (2652)
GGACGAGCAGAACGGCGACGTGCGCCGCACCGACGAGGTGGCCATCCACCAG polopt_hml (2652)

GAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCT polwt_hml (2704)
GAGGGCCGCGCCGCCAACCTGGGCACCACCAAGGAGGCCGACGCCGTGAGCT polopt_hml (2704)

ACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGC polwt_hml (2756)
ACAAGATCAGCCGCGAGCACAAGGGCGACACCAACCCCCGCGAGTACGCCGC polopt_hml (2756)

TTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGG polwt_hml (2808)
CTGCAGCCTGGACGACTGCATCAACGGCGGCAAGAGCCCCTACGCCTGCCGC polopt_hml (2808)

AGCAGCTGCAGC---TAA                                    polwt_hml (2860)
AGCAGCTGCAGCGCTTAA                                    polopt_hml (2860)
```

Env manipulation:

Start with SEQ ID 81 (SEQ ID 83); manipulate to SEQ ID 82:

```
envwt_HML2    ATGAACCCAAGCGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGA
envopt_HML2   ATGAACCCCAGCGAGATGCAGCGCAAGGCCCCCCCCCCGCCGCCGCCGCCACCGCAACCGC

envwt_HML2    GCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCA
envopt_HML2   GCCCCCCTGACCCACAAGATGAACAAGATGGTGACCAGCGAGGAGCAGATGAAGCTGCCC

envwt_HML2    TCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCT
envopt_HML2   AGCACCAAGAAGGCCGAGCCCCCCACCTGGGCCCAGCTGAAGAAGCTGACCCAGCTGGCC

envwt_HML2    ACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCC
envopt_HML2   ACCAAGTACCTGGAGAACACCAAGGTGACCCAGACCCCCGAGAGCATGCTGCTGGCCGCC

envwt_HML2    TTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTAAC
envopt_HML2   CTGATGATCGTGAGCATGGTGGTGAGCCTGCCCATGCCCGCCGGCGCCGCCGCCGCCAAC

envwt_HML2    TATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGGAT
envopt_HML2   TACACCTACTGGGCCTACGTGCCCTTCCCCCCCCTGATCCGCGCCGTGACCTGGATGGAC

envwt_HML2    AATCCTACAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCATAGATGATCGC
envopt_HML2   AACCCCACCGAGGTGTACGTGAACGACAGCGTGTGGGTGCCCGGCCCCATCGACGACCGC
```

```
envwt_HML2    TGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCATTAT
envopt_HML2   TGCCCCGCCAAGCCCGAGGAGGAGGGCATGATGATCAACATCAGCATCGGCTACCACTAC

envwt_HML2    CCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTG
envopt_HML2   CCCCCCATCTGCCTGGGCCGCGCCCCCGGCTGCCTGATGCCCGCCGTGCAGAACTGGCTG

envwt_HML2    GTAGAAGTACCTACTGTCAGTCCCATCTGTAGATTCACTTATCACATGGTAAGCGGGATG
envopt_HML2   GTGGAGGTGCCCACCGTGAGCCCCATCTGCCGCTTCACCTACCACATGGTGAGCGGCATG

envwt_HML2    TCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAATTT
envopt_HML2   AGCCTGCGCCCCCGCGTGAACTACCTGCAGGACTTCAGCTACCAGCGCAGCCTGAAGTTC

envwt_HML2    AGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTT
envopt_HML2   CGCCCCAAGGGCAAGCCCTGCCCCAAGGAGATCCCCAAGGAGAGCAAGAACACCGAGGTG

envwt_HML2    TTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGA
envopt_HML2   CTGGTGTGGGAGGAGTGCGTGGCCAACAGCGCCGTGATCCTGCAGAACAACGAGTTCGGC

envwt_HML2    ACTATTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTCAG
envopt_HML2   ACCATCATCGACTGGGCCCCCCGCGGCCAGTTCTACCACAACTGCAGCGGCCAGACCCAG

envwt_HML2    TCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAGAC
envopt_HML2   AGCTGCCCCAGCGCCCAGGTGAGCCCCGCCGTGGACAGCGACCTGACCGAGAGCCTGGAC

envwt_HML2    AAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATC
envopt_HML2   AAGCACAAGCACAAGAAGCTGCAGAGCTTCTACCCCTGGGAGTGGGGCGAGAAGGGCATC

envwt_HML2    TCTACCCCAAGACCAAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGG
envopt_HML2   AGCACCCCCCGCCCCAAGATCGTGAGCCCCGTGAGCGGCCCCGAGCACCCCGAGCTGTGG

envwt_HML2    AGGCTTACTGTGGCTTCACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAACA
envopt_HML2   CGCCTGACCGTGGCCAGCCACCACATCCGCATCTGGAGCGGCAACCAGACCCTGGAGACC

envwt_HML2    AGAGATCGTAAGCCATTTTATACTATTGACCTGAATTCCAGTCTAACAGTTCCTTTACAA
envopt_HML2   CGCGACCGCAAGCCCTTCTACACCATCGACCTGAACAGCAGCCTGACCGTGCCCCTGCAG

envwt_HML2    AGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCC
envopt_HML2   AGCTGCGTGAAGCCCCCCTACATGCTGGTGGTGGGCAACATCGTGATCAAGCCCGACAGC

envwt_HML2    CAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGG
envopt_HML2   CAGACCATCACCTGCGAGAACTGCCGCCTGCTGACCTGCATCGACAGCACCTTCAACTGG

envwt_HML2    CAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGAC
envopt_HML2   CAGCACCGCATCCTGCTGGTGCGCGCCCGCGAGGGCGTGTGGATCCCCGTGAGCATGGAC

envwt_HML2    CGACCGTGGGAGGCCTCGCCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTGTTTTA
envopt_HML2   CGCCCCTGGGAGGCCAGCCCCAGCGTGCACATCCTGACCGAGGTGCTGAAGGGCGTGCTG

envwt_HML2    AATAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTC
envopt_HML2   AACCGCAGCAAGCGCTTCATCTTCACCCTGATCGCCGTGATCATGGGCCTGATCGCCGTG

envwt_HML2    ACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTTT
envopt_HML2   ACCGCCACCGCCGCCGTGGCCGGCGTGGCCCTGCACAGCAGCGTGCAGAGCGTGAACTTC

envwt_HML2    GTTAATGATTGGCAAAAAAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAA
envopt_HML2   GTGAACGACTGGCAGAAGAACAGCACCCGCCTGTGGAACAGCCAGAGCAGCATCGACCAG

envwt_HML2    AAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTC
envopt_HML2   AAGCTGGCCAACCAGATCAACGACCTGCGCCAGACCGTGATCTGGATGGGCGACCGCCTG

envwt_HML2    ATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTTGTATT
```

```
envopt_HML2    ATGAGCCTGGAGCACCGCTTCCAGCTGCAGTGCGACTGGAACACCAGCGACTTCTGCATC

envwt_HML2     ACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAG
envopt_HML2    ACCCCCCAGATCTACAACGAGAGCGAGCACCACTGGGACATGGTGCGCCGCCACCTGCAG

envwt_HML2     GGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCA
envopt_HML2    GGCCGCGAGGACAACCTGACCCTGGACATCAGCAAGCTGAAGGAGCAGATCTTCGAGGCC

envwt_HML2     TCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGCAGGAGTTGCTGATGGC
envopt_HML2    AGCAAGGCCCACCTGAACCTGGTGCCCGGCACCGAGGCCATCGCCGGCGTGGCCGACGGC

envwt_HML2     CTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACTACGATTATAAAT
envopt_HML2    CTGGCCAACCTGAACCCCGTGACCTGGGTGAAGACCATCGGCAGCACCACCATCATCAAC

envwt_HML2     CTCATATTAATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAG
envopt_HML2    CTGATCCTGATCCTGGTGTGCCTGTTCTGCCTGCTGCTGGTGTGCCGCTGCACCCAGCAG

envwt_HML2     CTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGACGATGGCGGTTTTGTCGAAA
envopt_HML2    CTGCGCCGCGACAGCGACCACCGCGAGCGCGCCATGATGACCATGGCCGTGCTGAGCAAG

envwt_HML2     AGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGGCCTAA
envopt_HML2    CGCAAGGGCGGCAACGTGGGCAAGAGCAAGCGCGACCAGATCGTGACCGTGAGCGTGGCCTAA
```

IN VITRO EXPRESSION OF GAG SEQUENCES

Three different gag-encoding sequences were cloned into the pCMVKm2 vector:

(1) gag opt HML-2 (SEQ ID 54, including SEQ ID 62 and encoding SEQ ID 70 – Fig. 5).

(2) gag opt PCAV (SEQ ID 80, including SEQ ID 77 and encoding SEQ ID 79 – Fig. 8).

(3) gag wt PCAV (SEQ ID 53, including SEQ ID 76 and encoding SEQ ID 78 – Fig. 4).

The vectors were used to transfect 293 cells in duplicate in 6-well plates, using the polyamine reagent *TransIt™ LT-1* (PanVera Corp, Madison WI) plus 2 μg DNA.

Cells were lysed after 48 hours and analyzed by western blot using pooled mouse antibody against HML2-gag as the primary antibody (1:400), and goat anti-mouse HRP as the secondary antibody (1:20000). Figure 10 shows that 'gag opt PCAV' (lane 2) expressed much more efficiently than 'gag wt PCAV' (lane 3). Lane 1 ('gag opt HML-2') is more strongly stained than lane 2 ('gag opt PCAV'), but this could be due to the fact that the primary antibody was raised against the homologous HML-2 protein, rather than reflecting a difference in expression efficiency. To address this question, antibodies were also raised against the PCAV product and were used for Western blotting. Figure 11A shows results using the anti-HML2 as the primary antibody (1:500), and Figure 11B shows the results with anti-PCAV (1:500). Each antibody stains the homologous protein more strongly than the heterologous protein.

NUCLEIC ACID IMMUNIZATION

Vectors of the invention are purified from bacteria and used to immunize mice.

T CELL RESPONSES TO PCAV GAG

CB6F1 mice were intramuscularly immunized with pCMVKm2 vectors encoding PCAV gag (Figures 4 & 8) and induction of gag-specific CD4+ and CD8+ cells were measured.

Mice received four injections of 50μg plasmid at week 0, 2, 4 and 6. These plasmids included the wild type gag sequence (SEQ ID 76). Mice were then split into two separate groups for further work.

The first group of three mice received a further 50μg of plasmid at 25 weeks, but this plasmid included the optimized gag sequence (SEQ ID 77). Eleven days later spleens were harvested and pooled and a single cell suspension was prepared for culture. Spleen cells (1×10^6 per culture) were cultured overnight at 37°C in the absence ("unstimulated") or presence ("stimulated") of 1×10^7 plaque-forming units (pfu) of a recombinant vaccinia which contains the PCAV gag sequence ("rVV-gag", produced by homologous recombination of cloning vector pSC11 [116], followed by plaque purification of recombinant rVVgag). Duplicate stimulated and unstimulated cultures were prepared. The following day Brefeldin A was added to block cytokine secretion and cultures were continued for 2 hours. Cultures were then harvested and stained with fluorescently-labeled monoclonal antibodies for cell surface CD8 and intracellular gamma interferon (IFN-γ). Stained samples were analyzed by flow cytometry and the fraction of CD8+ cells that stained positively for intracellular IFN-γ was determined. Results were as follows:

Culture condition	Culture #1	Culture #2	Average
Unstimulated	0.10	0.14	0.12
Stimulated	1.51	1.27	1.39
		Difference	1.27

An average of 1.27% of the pooled splenic CD8+ cells synthesized IFN-γ in response to stimulation with rVV-gag. This demonstrates that the DNA immunization induced CD8+ T cells that specifically recognized and responded to PCAV gag.

The second group of four mice received a further 50μg of plasmid at 28 weeks, but this plasmid included the optimized gag sequence (SEQ ID 77). Twelve days later spleens were harvested. As a specificity control, a spleen was also obtained from a CB6F1 mouse that had been vaccinated with a pCMV-KM2 vector encoding HML2 env.

Single cell suspensions from individual spleens were prepared for culture. Spleen cells (1×10^6 per culture) were cultured overnight at 37°C in the absence of stimulation or in the presence of 1×10^7 pfu rVV-gag. As a specificity control, additional cultures contained another recombinant vaccinia virus, rVV-HIVgp160env.SF162 ("rVV-HIVenv" – contains full-length

env gene from SF162 isolate of HIV-1), which was not expected to cross-react with either gag or env from PCAV.

Duplicate cultures were prepared for each condition. The following day Brefeldin A was added to block cytokine secretion and anti-CD28 antibody was added to co-stimulate CD4 T cells. Cultures were continued for 2 hours and then harvested and stained with fluorescently-labeled monoclonal antibodies for cell surface CD8 and CD4 and intracellular IFN-γ. Stained samples were analyzed by flow cytometry and the fractions of CD8+CD4- and CD4+8- T cells that stained positively for intracellular IFN-γ were determined. Results are shown in the following table, expressed as the % of stained cells in response to stimulation by either PCAV gag or HIV env during spleen culture, after subtraction of the average value seen with cells which were not stimulated during spleen culture:

Spleen culture stimulation	**Vector administered at 28 weeks**				
	PCAV gag	PCAV gag	PCAV gag	PCAV gag	PCAV env
CD8					
PCAV gag	1.32	1.88	3.00	2.09	0.13
HIV env	0.04	0.12	-0.02	0.23	0.05
CD4					
PCAV gag	0.26	0.17	0.40	0.22	-0.01
HIV env	0.01	-0.02	-0.03	0.01	-0.02

For the 4 mice that had been vaccinated with a vector encoding PCAV gag, therefore, the rVV-gag vector stimulated 1.32% to 3.00% of CD8+ T cells to produce IFN-γ. However, there were few CD8+ T cells (<0.23%) that responded to the irrelevant rVV-HIVgp160env vector. The CD8+ T cell response is thus specific to PCAV gag. Furthermore, the control mouse that was immunized with PCAV env had very few CD8+ T cells (0.13%) which responded to the vaccinia stimulation.

Similarly, vaccination with PCAV gag, but not with PCAV env, induced CD4+ T cells specific for PCAV gag (0.17% to 0.40%).

DNA immunization with vectors encoding PCAV gag thus induces CD8+ and CD4+ T cells that specifically recognize and respond to the PCAV gag antigen.

VIRUS-LIKE PARTICLES

293 cells were fixed 48 hours after transient transfection with pCMV-gag, either from HML-2 or from PCAV, and inspected by electron microscopy (Figure 12). VLPs were produced in both cases, but these were mainly intracellular for PCAV and mainly secreted for HML-2.

The assembly of viable VLPs from PCAV and HML-2 indicates that the gag protein has retained its essential activity even though the endogenous virus is "dormant" and might thus be expected to be subject to mutational inactivation.

The above description of preferred embodiments of the invention has been presented by way of illustration and example for purposes of clarity and understanding. It is not intended to be exhaustive or to limit the invention to the precise forms disclosed. It will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that many changes and modifications may be made thereto without departing from the spirit of the invention. It is intended that the scope of the invention be defined by the appended claims and their equivalents.

SEQUENCE LISTING INDEX

SEQ ID	DESCRIPTION
1-9	Gag sequences
10-14	Prt sequences
15-21	Pol sequences
22-28	Env sequences
29-31	cORF sequences
32-37	PCAP sequences
38-50	Splice variants A-M sequences
51	pCMVKm2.cORFopt HML-2 (Figure 2)
52	pCMVKm2.pCAP5opt HML-2 (Figure 3)
53	pCMVKm2.gag wt PCAV (Figure 4)
54	pCMVKm2.gagopt HML-2 (Figure 5)
55	pCMVKm2.Protopt HML-2 (Figure 6)
56	pCMVKm2.Polopt HML-2 (Figure 7)
57-66	Nucleotide sequences pre- and post-manipulation
67	Manipulated cORF
68	Manipulated PCAP5
69 & 70	Gag — pre- and post-manipulation
71 & 72	Prt — pre- and post-manipulation
73 & 74	Pol — pre- and post-manipulation
75	PCAV, from the beginning of its first 5' LTR to the end of its fragmented 3' LTR
76 & 77	PCAV Gag nucleotide sequences — pre-and post manipulation
78 & 79	PCAV Gag amino acid sequences — pre-and post manipulation
80	pCMVKm2.gagopt PCAV (Figure 8)
81	Wild-type env from HML-2
82	Optimized env from HML-2
83	Amino acid sequence encoded by SEQ IDs 81 & 82

NB:

- SEQ IDs 1 to 9 disclosed in reference 1 as SEQ IDs 85, 91, 97, 102, 92, 98, 103, 104 & 146
- SEQ IDs 10 to 14 disclosed in reference 1 as SEQ IDs 86, 99, 105, 106 & 147
- SEQ IDs 15 to 21 disclosed in reference 1 as SEQ IDs 87, 93, 100, 107, 94, 108 & 148
- SEQ IDs 22 to 28 disclosed in reference 1 as SEQ IDs 88, 95, 101, 107, 96, 108 & 149
- SEQ IDs 29 to 31 disclosed in reference 1 as SEQ IDs 89, 90 & 109
- SEQ IDs 32 to 37 disclosed in reference 1 as SEQ IDs 10, 11, 12, 7, 8 & 9
- SEQ IDs 38 to 50 disclosed in reference 1 as SEQ IDs 28-37, 39, 41 & 43
- SEQ ID 75 disclosed in reference 3 as SEQ ID 1.

REFERENCES (the contents of which are hereby incorporated in full by reference)

1. International patent application WO02/46477 (PCT/US01/47824).
2. US patent application 10/016,604 (filed December 7, 2001).
3. International patent application PCT/US02/39136 (filed December 9, 2002).
4. Andersson *et al.* (1999) *J. Gen. Virol.* 80:255-260.
5. Mayer *et al.* (1999) *Nat. Genet.* 21 (3), 257-258 (1999)
6. Ono *et al.*, (1986) *J. Virol.* 60:589
7. US patent 5,858,723
8. Barbulescu *et al.* (1999) *Curr. Biol.* 9:861-868.
9. Zsíros *et al.* (1998) *J. Gen. Virol.* 79:61-70.
10. Tönjes *et al.* (1999) *J. Virol.* 73:9187-9195.
11. Genbank accession number AB047240
12. Paces *et al.* (2002) *Nucleic Acids Res.* 30:205-206.
13. Reus *et al.* (2001) *Genomics* 72:314-320.
14. Dunham *et al.* (1999) *Nature* 402:489-495.
15. Löwer *et al.* (1996) *Proc. Natl. Acad. Sci USA* 93:5177
16. Boese *et al.* (2000) *Oncogene* 19:4328-4336.
17. Mueller- Lantzsch *et al.*, AIDS Research and Human Retroviruses 9:343-350 (1993)
18. Berkhout *et al.* (1999) *J. Virol.* 73:2365-2375.
19. Löwer *et al.* (1995) *J. Virol.* 69:141-149.
20. Magin *et al.* (1999) *J. Virol.* 73:9496-9507.
21. Magin-Lachmann (2001) *J. Virol.* 75(21):10359-71.
22. International patent application PCT/US02/39344 (filed December 9, 2002).
23. Geysen *et al.* (1984) *PNAS USA* 81:3998-4002.
24. Carter (1994) *Methods Mol Biol* 36:207-23.
25. Jameson, BA *et al.*, 1988, *CABIOS* 4(1):181-186.
26. Raddrizzani & Hammer (2000) *Brief Bioinform* 1(2):179-89.
27. De Lalla *et al.* (1999) *J. Immunol.* 163:1725-29.
28. Brusic *et al.* (1998) *Bioinformatics* 14(2):121-30
29. Meister *et al.* (1995) *Vaccine* 13(6):581-91.
30. Roberts *et al.* (1996) *AIDS Res Hum Retroviruses* 12(7):593-610.
31. Maksyutov & Zagrebelnaya (1993) *Comput Appl Biosci* 9(3):291-7.
32. Feller & de la Cruz (1991) *Nature* 349(6311):720-1.
33. Hopp (1993) *Peptide Research* 6:183-190.
34. Welling *et al.* (1985) *FEBS Lett.* 188:215-218.
35. Davenport *et al.* (1995) *Immunogenetics* 42:392-297.
36. US patent 5,858,723
37. Johnston *et al.* (2001) *Ann Neurol* 50(4):434-42.
38. Medstrand *et al.* (1998) *J Virol* 72(12):9782-7.
39. Gennaro (2000) *Remington: The Science and Practice of Pharmacy.* 20th edition, ISBN: 0683306472.

40. WO 93/14778
41. Findeis *et al.*, *Trends Biotechnol.* (1993) *11*:202
42. Chiou *et al.* (1994) *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer.* ed. Wolff
43. Wu *et al.*, *J. Biol. Chem.* (1988) *263*:621
44. Wu *et al.*, *J. Biol. Chem.* (1994) *269*:542
45. Zenke *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1990) *87*:3655
46. Wu *et al.*, *J. Biol. Chem.* (1991) *266*:338
47. Jolly, *Cancer Gene Therapy* (1994) *1*:51
48. Kimura, *Human Gene Therapy* (1994) *5*:845
49. Connelly, *Human Gene Therapy* (1995) *1*:185
50. Kaplitt, *Nature Genetics* (1994) *6*:148
51. WO 90/07936
52. WO 94/03622
53. WO 93/25698
54. WO 93/25234
55. US patent 5,219,740
56. WO 93/11230
57. WO 93/10218
58. US patent 4,777,127
59. GB Patent No. 2,200,651
60. EP-A- 0 345 242
61. WO 91/02805
62. WO 94/12649
63. WO 93/03769
64. WO 93/19191
65. WO 94/28938
66. WO 95/11984
67. WO 95/00655
68. Curiel, *Hum. Gene Ther.* (1992) *3*:147
69. Wu, *J. Biol. Chem.* (1989) *264*:16985
70. US patent 5,814,482
71. WO 95/07994
72. WO 96/17072
73. WO 95/30763
74. WO 97/42338
75. WO 90/11092
76. US patent 5,580,859
77. US patent 5,422,120
78. WO 95/13796
79. WO 94/23697

80. WO 91/14445
81. EP 0524968
82. Philip, *Mol. Cell Biol.* (1994) *14*:2411
83. Woffendin, *Proc. Natl. Acad. Sci.* (1994) *91*:11581
84. US patent 5,206,152
85. WO 92/11033
86. US patent 5,149,655
87. WO 92/11033
88. Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648.
89. Strugnell *et al.* (1997) *Immunol Cell Biol* 75(4):364-369.
90. Robinson & Torres (1997) *Seminars in Immunol* 9:271-283.
91. Brunham *et al.* (2000) *J Infect Dis* 181 Suppl 3:S538-43.
92. Svanholm *et al.* (2000) *Scand J Immunol* 51(4):345-53.
93. *DNA Vaccination - Genetic Vaccination* (1998) eds. Koprowski *et al.* (ISBN 3540633928).
94. *Gene Vaccination : Theory and Practice* (1998) ed. Raz (ISBN 3540644288).
95. WO90/14837
96. *Vaccine Design: subunit and adjuvant approach* (1995) ed. Powell & Newman (ISBN 030644867X).
97. WO00/07621
98. GB-2220221
99. EP-A-0689454
100. EP-A-0835318
101. EP-A-0735898
102. EP-A-0761231
103. WO99/52549
104. WO01/21207
105. WO01/21152
106. WO00/62800
107. WO00/23105
108. WO99/11241
109. WO98/57659
110. WO93/13202.
111. Johnson *et al.* (1999) *Bioorg Med Chem Lett* 9:2273-2278.
112. International patent application WO00/50078.
113. Singh *et al.* (2001) *J. Cont. Rele.* 70:267-276.
114. *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30.
115. Smith and Waterman, Adv. Appl. Math. (1981) 2: 482-489.
116. Chakrabarti *et al.* (1985) *Mol Cell Biol* 5:3403-3409.